The Maintenance Constraint: How Resource Boundaries Shape Cognitive Availability

Lluis Eriksson
Independent Researcher
lluiseriksson@gmail.com

July 4, 2026 (v2; v1: December 2025) ai.viXra:2512.0084

Status note. This is the outermost, philosophy-of-mind paper of the coherence-maintenance series (ai.viXra:2512.0060/0061/0064/0070/0071/0072/ 0073/0081, all at v2, 2026). It contains no theorems of its own; its single load-bearing technical input is the maintenance inequality, imported as an *anchor*. **v2**: that inequality is cited in its corrected form — the incremental (extra-power) statement holds against thermodynamically efficient baselines, with an unconditional total-power / entropy-production floor underneath it, per the corrected companion papers [1, 2] (v2); v1 cited the unqualified extra-power form, whose companion-v1 formulation was later found vacuous. The philosophical argument is unaffected. **Non-claims:** no collapse mechanism, no Born-rule derivation, no claim about phenomenological consciousness or quantum coherence in brains.

Abstract

Cognitive systems are resource-limited, but "resource limitation" is often invoked without distinguishing one-shot costs (forming a representation) from sustained costs (keeping it usable under noise). We argue that the availability of internal state features for control, integration, and report is constrained by their *maintainability* under finite budgets. As a technical anchor we cite a companion preprint deriving an operational maintenance inequality in explicit thermodynamic control models: incremental maintenance power has a non-arbitrary lower bound tied to dynamical fragility (in its corrected form, against efficient baselines, with an unconditional entropy-production floor beneath). This motivates an operational cut: a feasibility boundary separating maintainable from unmaintainable state features. We develop an auditable bridge argument (maintainability → stability → availability), propose a neutrality-friendly principle of *maintenance-feasibility bias*, and show how it can be incorporated into active inference (the Free-Energy Principle) as a maintenance penalty or constraint, aligning secondarily with Global Workspace accounts of access stability. We address objections and offer falsifiable predictions for synthetic agents and neuromorphic systems, plus an explicitly exploratory psychophysics subsection framed in terms of reportability and stability rather than phenomenology. We do not propose collapse mechanisms, do not derive the Born rule, and make no claims about phenomenological consciousness.

1 Introduction

Many debates in philosophy of mind and cognitive neuroscience presuppose that internal variables can be treated as stable quantities available for downstream computation, control, and report. Yet biological and artificial agents operate under noise, drift, coupling, and finite budgets (metabolic power, bandwidth, control precision, compute). This raises a neglected question: which internal state features are feasible to keep *available* over task-relevant timescales?

A common move is to invoke "resource limits" in broad terms. While true, this is too coarse to constrain architectures. Our aim is to sharpen it into an operational constraint:

availability is limited by maintainability. Some state features may be representable in principle but too fragile or too costly to keep stably available under realistic budgets; this shapes which representations dominate behavior and report.

States vs. state features. We speak of *state features* rather than exact microstates: what must be stabilized is typically a task-relevant feature (a coarse-graining, a decoded variable, a functional manifold, a correlation pattern), not an entire microscopic configuration.

Two methodological commitments guide the paper. (1) We treat "availability" as a functional notion (access for downstream use), not a claim about phenomenology; this targets access-level cognition and report, not the hard problem [4, 5, 6]. (2) We keep the physics anchor separate from philosophy-of-mind claims: the anchor supplies an example of a nontrivial lower bound on maintenance cost in an explicit model; it does not imply that brains implement quantum coherence, nor that consciousness depends on quantum effects.

Non-claims (explicit). We do not propose collapse mechanisms, do not derive the Born rule, and make no claims about phenomenological consciousness. The term "cut" is used operationally — a feasibility boundary under resource constraints, not an ontological boundary in nature.

2 Existing approaches to physical constraints on cognition

That cognition is physically constrained is not new; what is often missing is a principled separation between constraints on *computation/ representation* and constraints on *maintenance* of state features over time under noise and coupling. Computational and informational limits (bounded rationality [13], resource-rational analysis [14], Marr's levels [12]) bound what can be *computed* or *encoded*. Metabolic-cost accounts [15, 16] bound the *energy* of signaling. The present proposal is orthogonal to both: it concerns the ongoing cost of *holding* a state feature usable, which neither a computational-complexity bound nor a per-spike energy budget directly captures.

3 Technical anchor: a maintenance inequality and an operational cut

We summarize the anchor at a level sufficient for the philosophical argument, without requiring the full physics proof [1, 2].

3.1 The anchor inequality (corrected form)

In an explicit thermodynamic control model (finite-dimensional system, Markovian uncontrolled dynamics, a work reservoir/battery interacting via energy-conserving global unitaries with a thermal environment at temperature T), one defines an incremental maintenance power $P_{\text{extra}}(\rho)$ isolating the steady cost of stabilizing certain state features at fixed populations. The anchor result is the explicitly qualified pair of lower bounds

$$P_{\min}(\rho) \;\geq\; k_{\mathrm{B}}T\,\sigma(\rho), \qquad P_{\text{inc}}^{\text{eff}}(\rho) \;\geq\; k_{\mathrm{B}}T\,\dot{C}_{\text{loss}}(\rho) \quad \text{in the efficient/free-baseline regimes,} \tag{1}$$

where $C(\rho)$ is a relative-entropy-based coherence functional, $\dot{C}_{\text{loss}}(\rho)$ its instantaneous loss rate under the uncontrolled dynamics, and $\sigma(\rho) \geq 0$ the target's entropy production rate. Here $P_{\text{inc}}^{\text{eff}}$ denotes the *corrected* incremental quantity: extra power relative to thermodynamically efficient population-maintenance baselines, or the free-baseline case where $P_{\text{extra}} = P_{\min}$. No unqualified paired-infimum statement is intended.

Provenance of (1) **(v2).** Both bounds are proved in the companion papers [2, 1] (v2): the first is unconditional; the second holds against thermodynamically efficient population-maintenance baselines and becomes assumption-free when the stabilized feature's "populations" are already stationary (the passive-stability case of §4.5). An earlier companion formulation defined the incremental cost by an unconstrained comparison and was vacuous; the corrected form is the one used here. The *philosophically important* feature is structural and survives the correction intact: maintenance has a non-arbitrary lower bound tied to dynamical fragility.

We emphasize: the anchor is a *benchmark case study* of maintainability costs, not a claim about neural implementation. The argument does not require that the brain maintains quantum coherence; the relevant import is the existence of principled maintenance lower bounds in explicit physical models.

3.2 The operational cut

Given a finite budget P_{avail}, define a feasibility boundary:

> *Operational cut:* the boundary between state features whose relevant maintenance power P_{maint} — total (P_{min}), incremental against an efficient baseline ($P^{\mathrm{eff}}_{\mathrm{inc}}$), or free-baseline, depending on the control task — lies below the available budget P_{avail}, and those for which it does not.

Using the task-appropriate P_{maint} rather than a bare P_{extra} avoids transporting the incremental notion into cognitive contexts where no clean "populations baseline" exists. This cut is operational and agent-relative — it depends on available resources and the uncontrolled dynamics — not an ontological division in nature. Its cognitive role emerges once maintainability is connected to stability and availability.

3.3 Methodological status

Our use of the anchor is methodological, not neurobiological. It provides an existence proof that, in explicit control models, holding certain features against uncontrolled dynamics entails non-arbitrary lower bounds tied to fragility. We do not infer that brains instantiate the anchor's microphysics; we use it to motivate a general stance: in any noisy substrate, availability of fine-grained internal variables over task-relevant horizons may require ongoing stabilization whose cost is constrained by the substrate's loss rates and coupling structure. The claim defended here concerns the architecture-shaping role of such feasibility boundaries, not the universality of any specific inequality.

4 From maintainability to cognitive availability

The central bridge is stated as an explicit, auditable argument. The claim is *not* that maintainability is consciousness; rather, maintainability constrains which internal state features can stably participate in cognition in the sense of being available for control, integration, and report.

4.1 Three notions

Maintainability (control-feasibility). A target feature is maintainable, relative to the uncontrolled dynamics and control primitives, if there exists an admissible strategy keeping the system in (or near) a functionally acceptable set for the requisite duration within budget. In the anchor model, maintainability is constrained by a lower bound on incremental maintenance power.

Stability (dynamical robustness). A feature is stable over a horizon if, under actual perturbations and couplings, it does not drift rapidly in ways that destroy its functional role. Stability concerns trajectory timescales, not instantaneous existence of an encoding.

Availability (cognitive-functional). A feature is available if it can be reliably used for downstream tasks: guiding action, updating beliefs, broadcasting to other subsystems, supporting report. Following the access/phenomenal distinction [4, 7], we treat availability as separable from phenomenology; this paper concerns availability and control, not qualia.

Maintenance resources (non-quantum reading). Measurable ongoing costs — metabolic power, control effort, communication bandwidth, precision/feedback gain — required to keep relevant variables within functional bounds.

4.2 Operational proxies

Stability proxies. Over a horizon T: (i) mean deviation from a functional set K, (ii) dwell-time in K before exit, (iii) decay time of an autocorrelation of a task-relevant variable.

Availability proxies. Downstream readout at latency τ: (i) decoding accuracy under noise, (ii) control-performance degradation when the variable is perturbed, (iii) mutual information between the variable and downstream actions within the relevant window. The maintenance constraint predicts systematic trade-offs between these proxies and measured maintenance expenditure under budget manipulations.

4.3 The dependency chain

P1 (Availability requires task-relevant persistence). To be available for control, integration, or report, a feature must persist on timescales comparable to downstream latencies; a microstate overwritten before it can influence downstream processes is not available in the relevant sense.

P2 (Persistence under noise often requires maintenance). Uncontrolled dynamics degrade fine-grained features toward default configurations; unless the target lies in a naturally stable manifold or is architecturally protected, persistence typically requires maintenance (feedback stabilization, metabolic expenditure, error correction, active control).

P3 (Maintenance can have unavoidable lower bounds). Under explicit thermodynamic bookkeeping, sustaining deviations from default relaxation can impose non-zero lower bounds on ongoing resource use. The anchor provides a paradigmatic case (quantum coherence) as an operational benchmark: in its corrected reading, $P^{\mathrm{eff}}_{\mathrm{inc}}(\rho) \geq k_{\mathrm{B}} T \, \dot{C}_{\mathrm{loss}}(\rho)$ against efficient baselines (unconditional total-power floor beneath), linking incremental maintenance power to intrinsic dynamical fragility.

C (Availability is resource-constrained). From P1–P3: the set of features available to an agent is constrained by its resource budget; some candidates exist as encodings in principle but are unavailable in practice because too costly to maintain under actual noise and coupling.

> **Maintenance-feasibility bias:** when multiple internal encodings serve comparable functional roles, finite-budget agents tend to adopt encodings whose stability is achievable within available maintenance resources.

What the conclusion does not say. Not that cognition is nothing but maintenance cost, nor that phenomenology is determined by the bound: it is a constraint on availability and stability of internal variables under finite budgets.

4.4 Loading vs. holding: work vs. power

A central operational distinction separates *loading* a representation (one-shot shaping cost to enter a target state) from *holding* it (sustained cost to keep it from drifting). This disentangles "the system can encode X" from "the system can stably use X". An encoding may be cheap to load but expensive to hold, or vice versa, predicting dissociations: transient high-fidelity microstates unavailable for report, versus lower-fidelity states that dominate behavior because they persist. (In the anchor, this is exactly the one-shot-work vs maintenance-power distinction made precise in [2, 3] v2.)

4.5 Passive stability as a special case

Not all stable representations require active maintenance: strong attractors, passive memory traces, and structural constraints yield persistence without ongoing work. This is a special case where the uncontrolled dynamics already keep the state near target, so incremental maintenance cost is negligible — in the anchor model, a vanishing loss rate for the relevant feature, making the bound trivial. Notably, this is *precisely* the free-baseline case in which the corrected inequality (1) holds without further assumptions [2] (v2): the correction and the passive-stability limit are the same statement seen from two sides. The maintenance-feasibility bias reduces to the correct special case: availability is unconstrained by maintenance when dynamics are already aligned.

5 Connections to cognitive frameworks

5.1 Active inference / FEP with a maintenance penalty

Active inference proposes that agents minimize variational free energy as a proxy for surprise [10, 11]. The maintenance constraint adds an implementational dimension: even if a representational state reduces variational objectives, it may be too fragile to hold under realistic perturbations. This motivates a multi-objective trade-off

$$\min_{q,\pi} \mathbb{E}_{\pi}\big[F_{\mathrm{var}}(q)\big] + \lambda\, P_{\mathrm{maint}}(q,\pi), \qquad \lambda > 0, \tag{2}$$

with P_{maint} the task-appropriate maintenance power of §3.2 (in physical instances, the corrected efficient-/free-baseline $P^{\mathrm{eff}}_{\mathrm{inc}}$): fragile encodings that lower F_{var} either raise measured maintenance power or are abandoned once the budget is capped.

5.2 Global Workspace: access stability

Global Workspace accounts [9, 7, 8] describe broadcast/access dynamics for reportable content. The maintenance constraint is a secondary alignment: broadcast requires that content remain stable long enough to be accessed, so access stability is a maintainability condition on workspace occupancy, not a claim about the workspace's phenomenal status.

6 Objections

"If the brain is classical, the quantum anchor is irrelevant." The anchor is not a claim about neural microphysics; it demonstrates in an explicit model that (a) maintenance can carry principled lower bounds, and (b) those bounds are a matter of physics, not engineering inefficiency. The philosophical constraint is substrate-general: any noisy substrate with nonzero loss rates for the relevant feature inherits a maintainability condition, whatever its microphysics.

"This is just metabolic cost renamed." No: metabolic-cost accounts bound signaling energy; the maintenance constraint bounds the ongoing cost of *holding a feature usable*, which can be nonzero even when signaling is cheap, and zero (passive stability) even when signaling is expensive.

"Maintainability is not availability." Correct — it is a *necessary* condition via the P1–P3 chain, not identical to it; the paper claims constraint, not identity.

7 Falsifiable implications

Synthetic/neuromorphic agents. Under explicit energy/compute budgets, policies achieving lower F_{var} via fragile encodings will either increase measured maintenance power or revert to robust/low-cost encodings once available power is capped — a directly testable budget-manipulation prediction. **Dissociation prediction.** Cheap-to-load, expensive-to-hold features should be underrepresented in report relative to their encodability, and the gap should widen as budgets tighten.

Exploratory psychophysics (framed as reportability/stability, not phenomenology). If report requires access stability, manipulations that raise the maintenance cost of a decoded feature (increased noise/coupling on its substrate) should reduce its reportability at fixed encodability, with the effect scaling with the feature's intrinsic loss rate. This subsection is explicitly exploratory and makes no phenomenological claim.

8 Conclusion

"Resource limitation" in cognition is sharpened here into a specific, operational constraint: availability is limited by maintainability, and maintainability can carry principled lower bounds under finite budgets. A companion technical preprint provides a benchmark instance in which maintenance has a non-arbitrary lower bound tied to dynamical fragility [1, 2] (v2). The resulting maintenance-feasibility bias is a neutrality-friendly, falsifiable organizing principle for which internal representations finite-budget agents can keep available — with the physics anchor kept firmly separate from, and not load-bearing for, any claim about phenomenology.

A Changes relative to version 1

1. **Anchor corrected, argument preserved.** v1 cited the maintenance inequality (1) as an unqualified extra-power bound (its reference [1], an unnumbered companion preprint). The companion's v1 extra-power definition was subsequently found vacuous (an infimum over unlinked strategy pairs equals $-\infty$); the corrected hierarchy [2, 1] (v2) reads (1) as an efficient-baseline incremental bound with an unconditional entropy-production floor beneath. This paper cites the corrected form. The philosophical bridge (P1–P3, maintenance-feasibility bias) is unchanged and, if anything, sharper: the passive-stability special case (§4.5) is now identified with the corrected inequality's free-baseline regime.

2. **Series references added.** v1 cited only the single anchor preprint; v2 cites the full v2 series and situates the anchor within it.

3. **Non-claims retained verbatim**; scope tightened where the anchor is invoked (§3.1, P3).

References

[1] L. Eriksson, *Operational Coherence Maintenance and the Quantum–Classical Boundary: Formal Definitions, Falsifiable Protocols, and an Outlook for Cognitive Systems*, ai.viXra:2512.0081, v2 (2026; v1: 2025). Companion repository, `papers/0081-quantum-classical-boundary`.

[2] L. Eriksson, *The Conditional Maintenance Work Theorem*, ai.viXra:2512.0061, v2 (2026; v1: 2025). Companion repository, `papers/0061-maintenance-work`.

[3] L. Eriksson, *The Heisenberg Cut as a Resource Boundary*, ai.viXra:2512.0064, v2 (2026; v1: 2025). Companion repository, `papers/0064-heisenberg-cut`.

[4] N. Block, *On a confusion about a function of consciousness*, Behav. Brain Sci. **18**, 227 (1995).

[5] D. J. Chalmers, *Facing up to the problem of consciousness*, J. Conscious. Stud. **2**, 200 (1995).

[6] D. J. Chalmers, *The Conscious Mind* (OUP, 1996).

[7] S. Dehaene and L. Naccache, *Towards a cognitive neuroscience of consciousness*, Cognition **79**, 1 (2001).

[8] S. Dehaene, *Consciousness and the Brain* (Viking, 2014).

[9] B. J. Baars, *A Cognitive Theory of Consciousness* (CUP, 1988).

[10] K. Friston, *The free-energy principle: a unified brain theory?*, Nat. Rev. Neurosci. **11**, 127 (2010).

[11] T. Parr and K. Friston, *Generalised free energy and active inference*, Biol. Cybern. **113**, 495 (2019).

[12] D. Marr, *Vision* (W. H. Freeman, 1982).

[13] H. A. Simon, *A behavioral model of rational choice*, Q. J. Econ. **69**, 99 (1955).

[14] F. Lieder and T. L. Griffiths, *Resource-rational analysis*, Behav. Brain Sci. **43**, e1 (2020).

[15] S. B. Laughlin et al., *The metabolic cost of neural information*, Nat. Neurosci. **1**, 36 (1998).

[16] D. Attwell and S. B. Laughlin, *An energy budget for signaling in the grey matter of the brain*, J. Cereb. Blood Flow Metab. **21**, 1133 (2001).

[17] G. Tononi, M. Boly, M. Massimini, and C. Koch, *Integrated information theory*, Nat. Rev. Neurosci. **17**, 450 (2016).